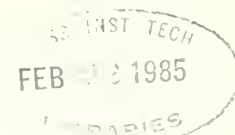


BASEMENT





WORKING PAPER
ALFRED P. SLOAN SCHOOL OF MANAGEMENT

The Retailer Facility Location Problem

by

Donald B. Rosenfield

January 1985

#1621-85

MASSACHUSETTS
INSTITUTE OF TECHNOLOGY
50 MEMORIAL DRIVE
CAMBRIDGE, MASSACHUSETTS 02139

The Retailer Facility Location Problem

by

Donald B. Rosenfield

January 1985

#1621-85

Visiting Associate Professor, Sloan School of Management

Arthur D. Little, Inc., Senior Consultant

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

ALL INFORMATION CONTAINED
HEREIN IS UNCLASSIFIED
DATE 02-22-1985 BY SP-6
HAW/EDD

I. Introduction

This paper discusses the problem of location of distribution centers in retailing networks and presents a successful case study for dealing with such a problem. While mathematical programming approaches for facility location problems for manufacturing firms have been extensively used [2], [3], [4], there has been very little treatment for the specific, somewhat simpler case, of locating facilities for a retailer. For a manufacturing firm for example, the problem of locating distribution centers (DC's) is compounded by manufacturing choice. That is, alternative plants can be used to produce different products, or customers can procure different products from different plants. In a retail environment, individual stores will generally have certain fixed requirements from each vendor. Hence, once a specific distribution center is assigned to a given store, all of the inbound (to the D.C., as based on fixed vendor requirements) and outbound (D.C. to stores) transportation and pipeline inventory costs and unit costs can be determined.

This determination of costs will in fact lead to a relatively simple subproblem given the specific facility locations. Specifically, a problem that is potentially three echelons can be reduced to two echelons. In most retailing environments, a store must be assigned uniquely to a distribution center. With the exception of this type of integer constraint, however, the assignment of stores to a fixed set of distribution centers is a simple, two-echelon transportation problem. We then consider a partial relaxation of the integer constraints by allowing more than one DC to

deliver merchandise to a single customer, but still requiring that each DC handle the same mix of vendors for that customer. With the partial relaxation, the simple transportation problem can be used as a basis of solution. In fact, since in practice stores are aggregated by zone, the assignment of multiple distribution centers to zones can be dealt with by splitting the individual stores to the distribution centers. Further as there are only a small number of distribution centers in most actual problems, the number of such situations will be small due to the optimality of basic solutions to linear programs. In sum, the simple transportation problem can be used as a basis for solving the retailer location problem.

A very fast algorithm for assigning stores to distribution centers based on this concept was developed and applied to a distribution center location problem for a major U.S. retailer that was expanding from one location to multiple locations.

The algorithm was based on the dual simplex method. The rationale for this approach is as follows: In any actual situation involving the assignment of stores or customers to distribution centers or warehouses, a good starting point is to assign each store to its least-cost distribution center. Since some facilities may be assigned more than their capacities, adjustments are required. In terms of the imbedded transportation problem, this initial solution is super-optimal but infeasible. But the same solution is feasible and suboptimal for the dual problem, and optimality can be obtained by a

1

2

0.05

1.8

3.000

0.05

0.05

0.05

0.05

0.050

0.05

0.05

0.05

0.05

0.05

0.05

0.05

0.05

sequence of dual pivots. As the number of stores may be large, but the number of distribution centers is small, the number of iterations required is small. In fact, if there is only one independent facility constraint, the dual-based approach is a linear algorithm and the algorithm is quite simple. One independent constraint will exist if there are only two distribution centers or a firm is contemplating expansion from one distribution centers to a network where the new distribution centers can have any capacities. In any case, for nearly all retailing chains in the U.S., the number of independent facility capacities will be small, and the dual-based approach should be very effective.

This type of procedure could be imbedded within a Bender's decomposition [3] or cross-decomposition scheme [5] to select locations from a set of locations. In the actual case study, since the number of total locations was two or three, optimal locations could be found by enumeration.

A second feature of the approach of this paper, in addition to the dual-based algorithm of the simplified mathematical program, was a flexible location system for potential facility locations. By locating network locations (demand points and vendor locations) using a coordinate system, facility locations can be specified by sets of coordinates, and system mileages can be automatically recalculated for each potential network. Hence by relating transportation costs and pipeline inventory costs (the capital charge of the inventory in the pipeline, which is proportional to the lead time) to the system distances, the relevant mathematical program and solution

corresponding to a given set of facility locations can be rapidly formulated and solved.

The remainder of this paper is organized as follows: In the next section, the retailer facility location is formulated, reflecting the simplified transportation problem obtained for a set of fixed facility locations and a partial relaxation of the unique assignment of facility to individual stores or store zones. We also describe why the partial relaxation maintains a simple form of the transportation problem but does not, in practice, yield infeasible solutions. In section 3, we describe the dual-based approach to the solution. In section 4, we describe the case study application of the approach and the flexible coordinate system approach for identifying facility locations. Finally, in section 5, we describe an analogous approach for locating warehouses in a manufacturing and distribution network. Again, by partial relaxation of the unique assignment of customers to warehouses, we still stipulate that any merchandise obtained by a customer from a warehouse be the correct mix of product requirements. The imbedded linear program is still very simple, and at least for the warehouse to customer part of the network, does not require separate treatment of products. The solution, as for the case of the retailer problem, is most likely still feasible.

II. Motivation and Formulation

To motivate the retailer facility location problem, we first describe the purpose of retailing distribution centers. Because of economies of scale in transportation costs, it becomes more economical to deliver goods in large loads to consolidation points,

100

101

102

103

104

105

106

107

108

109

110

111

112

113

114

115

116

117

118

119

120

121

122

123

124

125

126

127

128

129

and redistribute the loads from these points to customers or stores. While these distribution centers also act as storage locations, the major purpose is to capitalize on these economies of scale. In fact, in a typical retailing environment, inbound freight costs are generally two to three times as great, on a unit-weight basis, as outbound costs. The situation is depicted in Figure 1.

Transportation economies of scale also have a significant impact on the logistics structure of other industries. The automobile manufacturers, for example, use consolidation centers to handle the distribution of goods from vendors to plants, and also to distribute cars from plants to customers (using rail to ship to consolidation points and trucks to then ship to dealers). The distribution system of Federal Express Corporation, with its hub in Memphis Tennessee, is an example of effective capitalization of transportation economies of scale. Deliveries between any two locations are always routed through the hub. The U.S. postal service and the United Parcel Service also use distribution centers for the same purpose.

The key to the approach is the two-echelon formulation of the problem once a set of distribution centers have been selected. Denoting the variable x_{kj} as the amount of merchandise flowing from distribution center k to store (or store zone) j , we have a set of constraints for distribution center capacity, minimum store requirements, and the integer constraints that a store must receive all of its merchandise from a single distribution center. There is no need to define variables for an additional echelon as we can uniquely define all inbound and outbound costs. Outbound unit costs (from distribution center to stores) are the direct transportation and pipeline inventory unit costs. Inbound unit costs are the

846

8

846

846

846

846

846

846

846

846

846

846

846

846

846

846

846

846

846

846

846

846

846

846

846

average of the vendor to distribution center transportation and pipeline inventory unit costs corresponding to that store. For example, if a given store receives all of its merchandise from two vendors, then the inbound costs corresponding to a specific distribution center and that store are average the of the costs from the two vendors to the distribution center.

Note that this unit cost calculation is equivalent to simply totalling all of the inbound costs corresponding to the assignment of a distribution center to a store.

The requirement of dedicated distribution centers can indeed have a significant impact. While vendors near a store are best routed through a nearby distribution center, vendors far from the store might best be routed through a far distribution center. Hence the dedicated distribution center effectively eliminates the need for a third echelon.

The only other complication is that in some retailing networks, merchandise flowing from vendors in one part of the country to a distribution center in another part of the country may be transshipped through a closer distribution center. (See Figure 2.) However, the level of this transshipment activity is usually not limited. Hence formulation is usually a matter of defining the inbound unit costs as the minimum of direct shipment or transshipment through all other distribution centers.

The two-echelon problem with fixed distribution centers can thus be formulated as follows:

Let x_{kj} = Amount from DC k to store or store zone j

$$y_{kj} = \begin{cases} 1 & \text{if store } j \text{ assigned to DC } k \\ 0 & \text{otherwise} \end{cases}$$

C_{kj} = Cost from DC k to store j

$$= T_{kj} + I_{kj} + \sum_i a_{ij} S_{ik} / \sum_i a_{ij}$$

$$S_{ik} = \min_{k'} (T_{ik'} + I_{ik'} + T_{kk'}^* + I_{kk'}^*)$$

where

a_{ij} = Amount required from source i , for store or store zone j

$$B_j = \sum_i a_{ij}$$

A_k = Capacity at DC k

T_{kj} = Unit transit cost, k to j

I_{kj} = Unit pipeline inventory cost, k to j

S_{ik} = Total inbound unit cost, vendor i to DC k

$T_{kk'}^*, I_{kk'}^*$ = Tranship transit and pipeline inventory cost from DC k to DC k'

minimize $\sum C_{kj} x_{kj}$

such that $\sum_j x_{kj} \leq A_k \quad \forall k$

$$\sum_k x_{kj} \geq \sum_i a_{ij} = B_j \quad \forall j$$

$$x_{kj} \geq y_{kj} \sum_i a_{ij} \quad \forall j, k$$

$$\sum_k y_{kj} = 1$$

Notice that with the exception of the integer stipulation that a store receive all of its merchandise from a single distribution center, the optimization problem given a set of locations and

1937

1938

1939

1940

1941

1942

1943

1944

1945

1946

1947

1948

1949

1950

1951

1952

1953

1954

1955

1956

1957

1958

1959

capacities is a simple transportation problem. If we relax the integer constraints above, the formulation still requires that if a store receives less than 100 percent of its merchandise from a given distribution center, then the mix of merchandise from the various vendors is still preserved. That is, for example, if a store receives 50% from a given DC, then the goods from that DC comprise 50% of the requirements from each vendor. This is only a partial relaxation, and furthermore, since stores will be aggregated into store zones, the splits (relaxations) may correspond to the assignment of different stores within the zone to the different distribution centers. Also, there will be at most a small number of these situations, due to the optimality of basic solutions to the transportation problem. In summary, the linear program relaxation of the problem should not yield problems of infeasibility to the original problem.

The transportation problem relaxation was the problem solved in the computer package and case study. Stores were aggregated into zones and in the few instances where two distribution centers split a zone, the differences were assumed to correspond to stores. Thus solution of the problem consisted of

- a. posing locations
- b. solving the transportation problem
- c. going to step a

Step a could be an enumeration or a decomposition procedure. In any case, however, the transportation problem is the appropriate subproblem.

111

112

113

114

115

116

117

118

119

120

121

122

123

124

125

126

127

128

129

130

III. Dual-Based Solution to Linear Program

To solve the transportation subproblem, we used the dual-simplex algorithm. To motivate the use of the dual problem, note that a natural approach to finding a solution is to assign each zone to least-cost distribution center and then make adjustments. This is indeed a dual approach and the initial solution to the primal is generally superoptimal but infeasible. With a few distribution centers and many store zones, not many iterations are required. In fact, for two distribution centers, we can use the following greedy (technically semi-greedy in the sense that a variable can be changed once) heuristic that emulates the dual simplex method:

1. Assign each zone to least-cost DC.
2. If all DC's are within capacity, stop.
3. If either DC handles an amount that is above capacity, adjust zone whose cost difference between the other DC and the capacity-violating DC is minimal. The amount of the adjustment is the minimum of activity for that zone and the level of the infeasibility.
4. Go to 2.

Note that the greedy heuristic will yield the optimal solution. The optimality of this type of procedure for the $2 \times n$ transportation problem was originally proven in [1] (although the procedure was not originally identified as a dual-based procedure).

We also note that the analogous type of procedure is optimal for multiple distribution centers with only one independent facility constraint, that is, where there is only one facility with either maximum capacity and/or minimum requirements. In this case, the revised step 3 is:

3. If the single DC with maximum capacity and/or minimum requirement handles an amount of merchandise that is above capacity (below minimum), then adjust the zone whose cost difference between the best other DC and the single constrained DC is minimal. Amount of adjustment is the minimum of activity for that zone and the level of the infeasibility.

Note that the revised step 3 can be applied to the usual two-facility case above. In this case there is in effect only one independent facility constraint since the second capacity, given the total system demand, simply translates into a minimum requirement for the first facility. Hence either one of the two facilities can be denoted as the fixed facility for the revised step 3. Thus, with two facilities, the algorithm seeks to adjust for the single capacity that is being exceeded after the initial least-cost assignment. With multiple facilities, the algorithm seeks to adjust in one direction or another to make sure that the facility with fixed capacity (and minimum requirements) is handling the correct amount of merchandise.

A useful observation concerning the greedy heuristic is that it is precisely the dual-simplex method. (This suggests that the dual simplex method might be appropriate for multiple locations.) Consider the general problem with two distribution centers and an arbitrary number of store zones.

The primal problem can be stated as

$$\sum x_{1j} = A_1$$

$$\sum x_{2j} = A_2$$

and

$$x_{1j} + x_{2j} \geq B_j \text{ for each market}$$

$$\min \sum C_{1j}x_{1j} + \sum C_{2j} x_{2j}$$

The dual problem is

$$z_1 + y_j \leq C_{1j}$$

$$z_2 + y_j \leq C_{2j}$$

$$z_1, z_2 \text{ unrestricted}$$

$$\max A_1z_1 + A_2z_2 + \sum B_jy_j$$

To develop an initial dual solution corresponding to the least-cost assignments of DC's to zones, we see that for zones j where source 1 is lower-cost,

$$z_1 - z'_1 + y_j = C_{1j} \text{ by complementary slackness.}$$

Otherwise

$$z_2 - z'_2 + y_j = C_{2j}$$

where the primed variables are the non-positive parts of z_1 and z_2 .

Now consider an example. Let

$$C = \begin{matrix} & \begin{matrix} \text{zone 1} & \text{zone 2} \end{matrix} \\ \begin{matrix} \text{DC1} \\ \text{DC2} \end{matrix} & \begin{bmatrix} 1 & 3 \\ 2 & 5 \end{bmatrix} \end{matrix}$$

$$A = 1, 3$$

$$B = 2, 2$$

The greedy heuristic is as follows:

Assign both zones to DC 1.

Minimum difference is zone 1, so assign 2 unit to DC 2.

Next minimum is zone 2, so assign 1 more unit to DC 2.

The dual is

$$z_1 + y_1 \leq 1$$

$$z_1 + y_2 \leq 2$$

$$z_1 + y_2 \leq 3$$

$$z_2 + y_2 \leq 5$$

$$\text{Max } z_1 + 3z_2 + 2y_1 + 2y_2$$

The initial solution, by complementary slackness, is $y_1 = 1$, $y_2 = 3$, and the tableau is (before canonical form) is

z_1	z_1'	z_2	z_2'	y_1	y_2	s_{11}	s_{12}	s_{21}	s_{22}	
1	-1			1		1				1
		1	-1	1			1			2
1	-1				1			1		3
		1	-1		1				1	5
1	-1	3	-3	2	2					

In setting up the canonical form of the tableau, you tabulate pairwise row differences, which are the adjustment values (cost differences) in evaluating changes from one distribution center to another. If the initial solution is not primal feasible, zeroing out one of the y_i columns gives a positive reduced cost for one z_i . In the example, canonical form is as follows (with the differences in DC costs for the two zones labelled).

z_1	z_1'	z_2	z_2'	y_1	y_2	s_{11}	s_{12}	s_{21}	s_{22}	
1	-1			1		1				1 zone 1 dif
-1	1	1	-1			-1	1			1
1	-1				1			1		3
-1	1	1	-1					-1	1	2 zone 2 dif
-3	3	3	-3			-2		-2		

Then each successive iteration subtracts the previous adjustment (cost difference) from the others. Hence the leaving variable corresponds to the zone of the next largest adjustment.

While the heuristic is only optimal for the case of independent constraints for a single facility only, it indicates that a dual-based procedure might be appropriate for the general case. There are two reasons for this. First, the initial solution of assigning each store or store zone to its least cost distribution center is always feasible for the dual and is usually an excellent starting point for the solution. Second, while there may be more than two distribution centers, there are almost always only a small number of these distribution centers, and the dual-simplex method should not require a large number of iterations.

IV. A Case Study and an Interactive Computer Package

The author was recently involved in a case study for a major retailer undergoing an ambitious expansion program. The retailer was currently operating a single distribution center and was expanding its geographical base. The company was interested in examining the impacts of alternative capacities of its existing facility in any future expansion plans and was contemplating one or two new distribution centers. Hence the dual-based heuristic with multiple distribution centers but with only a single independent capacity was appropriate. The computer package developed by the author was used to examine a wide range of locations and capacity utilizations of the firms present facility.

To solve the problem, the author developed a computer package written in the language APL. APL is a high-level language with extensive matrix and vector algebra capabilities. It is very effective for developing prototype models for logistics analysis. For example, the dual-based heuristic described previously was written in a few lines of APL code.

Each line of code can perform system calculations. IF DCLOC and VLOC are $k \times 2$ and $m \times 2$ vectors giving latitudes and longitudes for distribution centers and vendor locations, then the statement DCLOC DIST VLOC, where DIST is a two-line code giving the great-circle formula, provides the $k \times m$ matrix of distances. Sales vectors can be multiplied by unit cost vectors (in single statements) to provide total costs. Matrix generators can be written very concisely. APL is also an interpretive language, and provides many additional operators (eg. matrix inverse) in addition to performing vector and matrix algebra.

The APL package incorporated two major features:

1. The dual-based heuristic described in this paper.
2. The capability of specifying and solving the merchandise flow for facilities located at any arbitrarily input locations.

The arbitrary specification of distribution center locations was based on a latitude and longitude coordinate system. Distances were calculated on the basis of the "great circle" formula.

Distances were converted into unit costs on the basis of the relationships between transportation costs and lead times and distances. Some of these relationships are depicted in figures 3 to

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

5. Lead time functions were designated for outbound shipments and for three different classes of vendors. There were similarly three classes of transportation functions as well as lead time and transportation functions for transshipment. Each of these functions were developed from a regression analysis of company supplied freight and lead time data.

Outbound freight costs could be fit very well to a linear regressions, as most outbound shipments were in full loads in company vehicles travelling to a set of stores. (Store zones were assigned adjustments to correspond to the correct number of store stops per delivery.) Inbound freight costs were somewhat more complex. Costs were variable and depended very much on the specific weight distribution for the particular vendor and destination combination. To deal with this complexity, the author analyzed the weight distribution and freight costs for vendors in each state of the country. Adjustment functions were derived for each vendor zone that were multiplied by the distance-based regression function to develop freight costs.

A demand data base was developed by examining the number of stores by sales volume in each of eighty geographic zones of the country and the vendor sales in each state of the country. As in many similar applications, the mass of each state was assumed to be located at the population centroid of that state. (U.S. population centroids are published by the U.S. census bureau.) The explosion of vendor to store zone demands (which were necessary to tabulate

inbound costs for each distribution center to store zone combination) was based on an independence assumption for vendors and store zones.

The algorithm for testing new distribution center locations was as follows:

1. Feed in capacities and locations of facilities.
2. Calculate distances using great circle fomula for all vendors (aggregated) to DC's and all DC's to store zones.
3. Calculate pipeline inventory and unit freight costs.
4. Solve LP.

As noted, the linear program could also be imbedded within a Bender's decomposition or cross-decomposition procedure to optimize merchandise flow for master problem specifications of facility locations. In our package, enumeration was sufficient.

The package identified the best distribution center locations for each stage of company growth, and the company is using specific locations identified by the computer package.

V. Extensions to Manufacturing Problem

As noted, the locations of distribution centers for a manufacturing company, even when the plants are fixed, is a more complex problem. As products can be selected from any of several plants, there is an extra degree of decision making. However the problem, given fixed distribution center locations, can still be somewhat simply formulated. In fact, while it is necessary to look at both the plant to warehouse (or distribution center) and warehouse to customer echelons, it is necessary to look at separate commodities

only in the plant to warehouse stage. The approach is similar to the retailing problem formulation. By recognizing (before relaxation) that a customer receives all its merchandise from a single warehouse, the plant to warehouse requirements can be based on warehouse to customer assignments. When the unique warehouse requirements are relaxed, the same product proportions can be maintained.

Specifically, given fixed locations, the problem is to

$$\text{minimize } \sum C_{kj}x_{kj} + E_{ikl}y_{ikl}$$

$$\text{where } \sum_j x_{kj} \leq A_k \quad \forall k$$

$$\sum_k x_{kj} \geq D_j \quad \forall j$$

$$\sum_i y_{ikl} \geq \sum_j x_{kj} d_{lj}/D_j \quad \forall k, l$$

$$\sum_k y_{ikl} \leq A_{il} \quad \forall i, l$$

where

x_{kj} = total product flow (in value or weight) for warehouse
k and customer zone j

y_{ikl} = total product flow of type l from plant i to warehouse k

C_{kj} , E_{ikl} are unit costs

D_j = demand at zone j

A_k = capacity of warehouse j

A_{il} = capacity of plant i, product l

d_{lj} = demand of product l at zone j

The third and fourth constraints are new. The third constraint states that each warehouse requires enough merchandise to meet the proportional requirements of each customer. The fourth constraint is plant capacity. While this problem is no longer a transportation problem, it is still a network problem and not very large. The dual based approach might still be useful, and the computer requirements to solve the problem for fixed locations are limited.

REFERENCES

1. Adolphson, D.L., and Thomas, G.N., "A Linear Time Algorithm for a $2 \times n$ Transportation Problem", SIAM Journal of Computing, Vol. 6, No. 3, Sept. 1, 1977.
2. Erlenkotter, "A Dual-Based Procedure for Uncapacitated Facility Location", Operations Research, November-December, 1978.
3. Geoffrion and Graves, "Multi-Commodity Distribution System Design by Benders Decomposition", Management Science, January, 1977.
4. Khumula, B.M., "An Efficient Branch and Bound Algorithm for the Warehouse Location Problem", Management Science, August, 1972.
5. Van Roy, T., "Cross-Decomposition for Mixed-Integer Programming Problems", Mathematical Programming, Vol. 25, 1983.

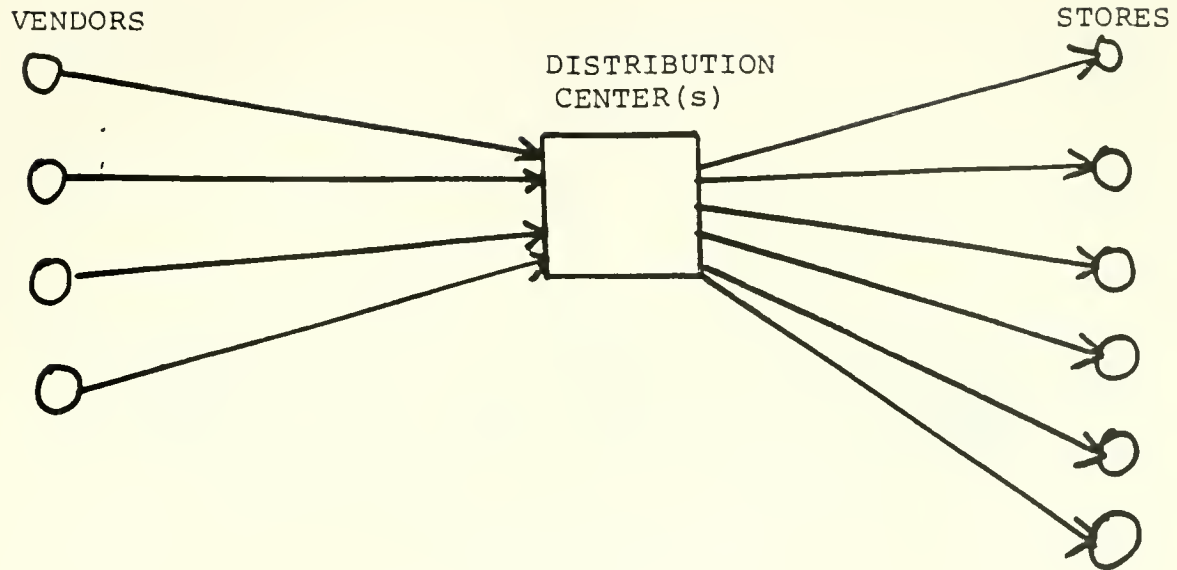


Figure 1: Consolidating Function

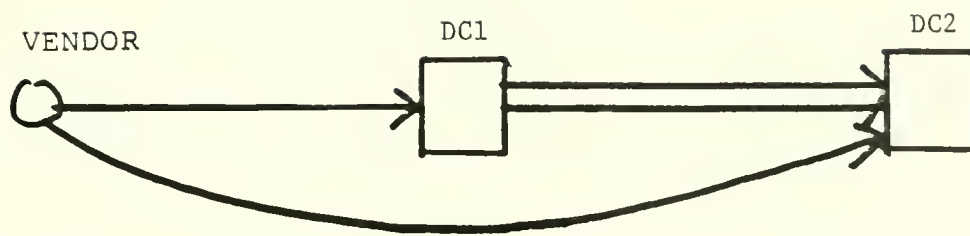
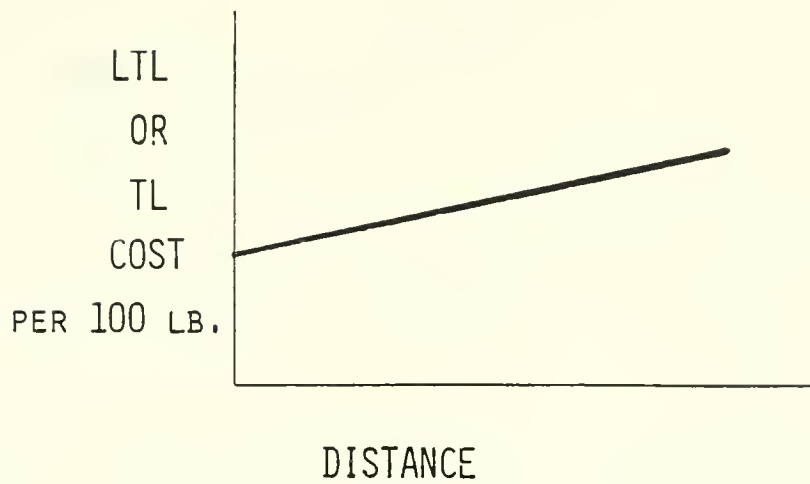
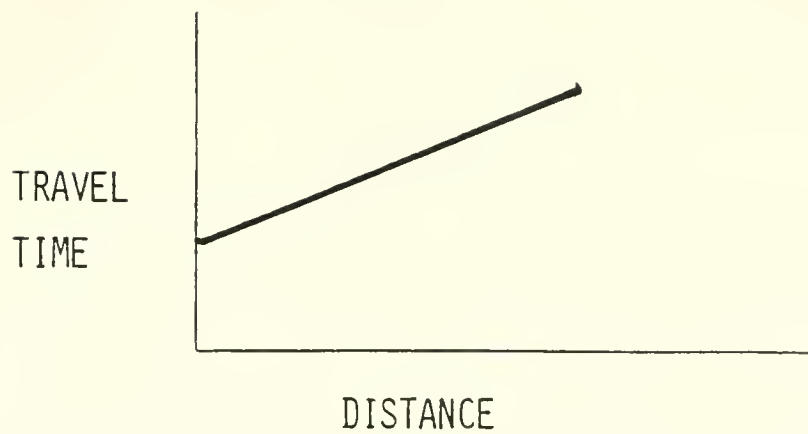


Figure 2: Transshipment option



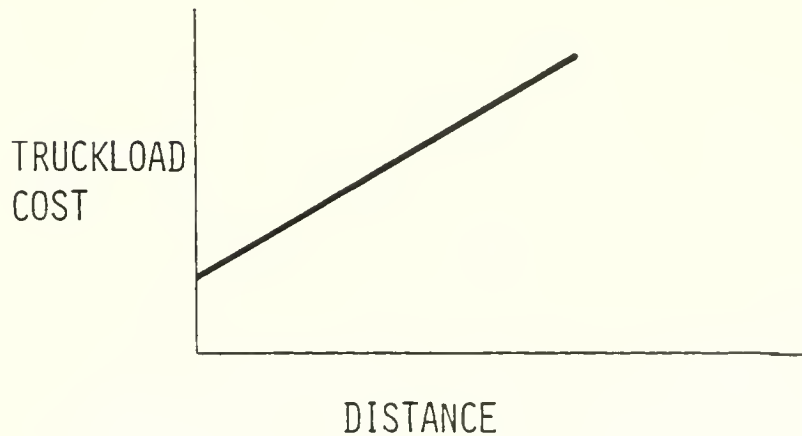
$$\text{UNIT COST} = \frac{\text{STANDARD LTL COST} \times \text{ZONE ADJUSTMENT}}{\text{VALUE PER 100 LB.}}$$

Figure 3: Inbound freight cost



$$\text{UNIT COST} = \text{DAYS TRAVEL TIME} \times \text{DAILY COST OF CAPITAL}$$

Figure 4: Lead time function



$$\text{UNIT COST} = \frac{(\text{TRUCKLOAD COST} + \text{STOPOFF CHANGE PER ZONE})}{\text{VALUE PER TRUCK}}$$

Figure 5: Outbound freight cost



UNIT COST = DAYS TRAVEL TIME X DAILY COST OF CAPITAL

UNIT COST = DAYS TRAVEL TIME X DAILY COST OF CAPITAL



UNIT COST = DAYS TRAVEL TIME X DAILY COST OF CAPITAL

UNIT COST = DAYS TRAVEL TIME X DAILY COST OF CAPITAL

MIT LIBRARIES



3 9080 004 481 831

Date Due BASEMENT

MAR 21 1961

APR 18 1961

MAY 2 1966

Lib-26-67

Bar
Code
On Last
Page

